DOCUMENT RESUME

ED 442 853 TM 031 273

AUTHOR Rozell, Diann S.
TITLE Contingency Table Statistics and Educational Reality: Problems with the Chi-Square Statistic.
PUB DATE 1999-11-00
NOTE 24p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Point Clear, AL, November 16-19, 1999).
PUB TYPE Information Analyses (070) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Chi Square; *Educational Research; Statistical Significance
IDENTIFIERS *Contingency Tables

ABSTRACT

The Pearson chi-square test can be useful in situations in which the researcher wishes to compare observed versus expected frequencies in categories, or cells, or a contingency table. Although these tests can be useful, various problems associated with their use and interpretation are common. First, the chi-square test is often the result of weak research questions. Second, chi-square tests are often the result of weak or erroneous information about data. An educational research data set is used to illustrate that statistically significant chi-squares often do not inform the researcher about the contributions of the cells in the contingency table, resulting in unclear conclusions or the use of additional statistical tests, neither of which is a promising alternative. (Contains 3 tables and 24 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

TM

Contingency Table Statistics and Educational Reality

ED 442 853

D. Rozell

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Contingency Table Statistics and Educational Reality: Problems With the Chi-Square Statistic

Diann S. Rozell

University of North Texas

Paper presented at the annual meeting of the Mid-South Educational Research Association, Point Clear, AL, November 16-19, 1999.

TM031273

BEST COPY AVAILABLE

Abstract

The Pearson chi-square test can be useful in situations in which the researcher wishes to compare observed versus expected frequencies in categories, or cells, of a contingency table. Although these tests can be useful, various problems associated with their use and interpretation are common. First, the chi-square test is often the result of weak research questions. Second, chi-square tests may yield weak or erroneous information about data. An educational research data set is used to illustrate that statistically significant chi-squares often do not inform the researcher about the contributions of the cells in the contingency table, resulting in unclear conclusions and/or utilization of additional statistical tests, neither of which is a promising alternative.

Contingency Table Statistics and Educational Reality: Problems With the Chi-Square Statistic

Chi-square, a nonparametric statistical test, compares the observed and expected frequency of occurrence of one or more nominal variables. It is often used when the research data are in the form of frequency counts. Karl Pearson's two dimensional, row by column (r by c) chi-square contingency tables have been available to social scientists since the development of the first inferential methods in 1900. Popham and Sirotnik in 1973 (p.284) argued that the chi-square test is "undoubtedly the most important member of the nonparametric family. In 1978, Mouly (p.199) suggested that the r by c test is "probably the best-known nonparametric test." Goodwin and Goodwin (1985) in a review of social science journals found the chi-square methodology employed in between eight and 17% of published articles.

The focal point of chi-square lies in the comparison of the observed frequencies of a given characteristic(s) or response(s) to the expected frequencies, and is represented by the χ^2. Observed frequencies (f_o) are the actual results observed in the data and are located within each of the categories or cells. The expected frequencies (f_e) are based on the theoretical number of observations that would fall into each category assuming some particular hypothesis. A common chi-square null hypothesis is

that an equal number of people should fall into each category, or, in other words, that no differences are expected in the frequencies (Couglin & Pagano, 1997). The computation of the chi-square statistical significance test, frequently considered a test of association or relationship between the two factors in a contingency table, is a relatively simple calculation. This computational simplicity may account for chi-square's abundance of use. Thompson (1988) provided a simple narrative for the chi-square calculations:

> For each of the k cells in the table, the difference between the observed and the expected count for the cell is squared and then divided by the cell's expected count. Each of these values is then summed across the number of cells to yield the calculated chi-square. For the chi-square tests of association, under an assumption that the null hypotheses is true, the expected values for the cells are computed by multiplying the column and row totals associated with each cell and then dividing the product by the number of entries in the total table. (p. 40)

The calculation of the degrees of freedom is equally as simple: the number of rows minus one times the number of columns minus one [(r - 1)(c-1)].

This paper reviews several common problems, noted as early as in 1949, related to the chi-square contingency table and

focuses on two specific problems currently encountered in educational research associated with the chi-square test of statistical significance.

Background

In 1900, when Pearson presented the chi-square test, he did not provide a limiting distribution test statistic. The value calculated for the test statistics was compared against existing tabled values. Over the next 30 to 40 years, individuals, including R. A Fisher, J. Neyman, E. S. Pearson and Karl Pearson himself, made contributions to both the theory and application of the chi-square test. This process culminated with a test statistic subsequently developed by Cramer in 1946 (Delucchi, 1981).

In 1949, Lewis and Burke addressed nine principal sources of error, they found regarding the use of chi-square (Delucchi, 1981):

1. Lack of independence among single events of measures
2. Small theoretical frequencies
3. Neglect of frequencies of non-occurrence
4. Failure to equalize the sum of the observed frequencies and the sum of theoretical frequencies
5. Indeterminate theoretical frequencies
6. Incorrect or questionable categorizing
7. Use of non-frequent data

8. Incorrect determination of the number of degrees of freedom
9. Incorrect computations

Deluchhi revisited this extensive article in 1981 due to his concern of continuing misuse and errors related to chi-square. He provided elaboration on several techniques of concern to educational researchers.

1. Partitioning
2. Log likelihood ratio
3. Correction for Continuity
4. Comparisons of two independent chi-squares
5. Analysis of ordered categories
6. Measures of association

Much discussion exists and considerable research was generated as to the appropriate expected size of the cell frequency. Fisher, in 1938, suggested that the cell frequency had to be greater than 5. Cramer, in 1946, suggested that the cell frequency had to be greater than 10. In 1952, Kendall suggested that the cell frequency had to be greater than 20 (Parshall & Kromrey, 1996). Each of these suggestions appears to have been overly conservative. Currently, the conservative rule of thumb, based on Cochran, is to avoid using the chi-square tests with expected cell frequencies less than 1 (obviously) or

when more than 20% of the contingency table cells have expected cell frequencies less than 5 (Prophet, 1999).

Problems related to chi-square encountered in educational research

The incorporation of chi-square analysis requires careful, logical examination of the prescribed study design. The chi-square is only a test of whether or not a null hypothesis of no association should be rejected. It is "not a measure of the degrees of relationship" (Best, 1981). This common misinterpretation of the chi-square test is not a problem with the test itself, but rather a misapplication or misconception of the statistical technique on the part of the researcher.

The use of chi-square permits the researcher to ascertain if two or more nominal variables are significantly related. Since the researcher has no scores to work with, the basic research question must address how individuals or items are distributed among various groups. In a chi-square analysis, the data are in the form of numbers of people or of items. Consequently, the investigator's questions cannot deal with how the mean scores of various groups of people may differ with respect to a particular variable or with the relationship between scores on two measures among a single group of people (Crowl, 1996).

Coughlin and Pagano (1997) offer guidelines to assist the researcher in determining methodology. Table 1 provides questions of methodology that a researcher might ask prior to initiating a study, along with responses that would lead to the selection of the chi-square contingency table as the appropriate analysis tool.

A basic assumption of the chi-square test of independence is that a subject contributes data to only one cell. Hence, the sum of all cell frequencies in the contingency table must be the same as the number of subjects in the experiment. Table 2 depicts the results of a hypothetical experiment in which each individual throws a ball into a basket once using his or her preferred hand and once using his or her non-preferred hand. The chi-square would be an invalid method to analyze these data considering that each individual contributed data to two cells. The total number of cell frequencies is 24, but the total number of subjects is 12.

Additionally the researcher should restrict the utilization of the chi-square test and, in turn, the research questions, to incidents in which the categories into which frequencies fall are discrete rather than continuous. A typical research study directs its attention to determine whether more boys or girls responded favorably to a particular form of math instruction. The gender of the child is a discrete variable (either boy or

girl). The math instruction is either experimental or non-experimental. Under these conditions, chi-square is the appropriate test of statistical significance.

Gall, Borg, and Gall (1996) have contented that the chi-square test is equally useful when the data or characteristics being considered are actually continuous variables that have been categorized. For example, in sociometric measurements, the achievement of a child is often a continuous variable. The researcher may use these continuous variables to categorize students into several groups such as "low-performing," "average," or "gifted" on the basis of the number of points each child receives. Arguments can be made that suggests that the sociometric category into which the student is classified provide a more meaningful basis for analyzing the data than the true achievement score. Because the categories of contingency tables are relatively limited, the researcher may consider increasing the expected values by increasing the sample size. Additional data are often difficult to obtain. The remaining option is to collapse columns and/or rows. This procedure can lead to a scenario where a failure to reject the null hypothesis for the collapsed table does not eliminate the possibility of non-independence in the original table, because collapsing can destroy evidence of non-independence (Prophet, 1999).

Several statisticians have contradicted this view. For example, Kerlinger (1986) contended that if the research's continuous dependent variables are converted to a nominal scale for the sake of comparison, the researcher would consciously and deliberately throw variance away. Additionally, when researchers regroup their data, the procedure effects the power of subsequent statistical tests (Timm, 1971). When numerical variables appear, they should be analyzed with a specific tool that exploits their numerical nature. Chi-square does not accurately accomplish this task. Hence, despite claims to the contrary, truncation of continuous variables into categories for the purpose of performing chi-squares is not an acceptable research practice.

Turning to another research situation, many educational studies focus on demographic characteristics and specific questionnaire responses. The appropriate statistical test could be a two-way chi-square; however, problems arise when "The investigator is able to explain the frequent small number of [statistically] significant results perfectly, although seldom have the [statistically] significant results been predicted a priori." (Stevens, 1996, p. 9).

Contributions of the contingency table cells in analyzing statistical significance

Even when chi-square tests are appropriately employed, the results of the tests are often misinterpreted. According to Thompson (1988), the chi-square tests a general null hypothesis and does not inform the researcher as to the nature of the relationship between factors included in the analysis. Specifically, a statistically significant result does not inform the research as to which cells generated the result. The problem is to determine at what point the actual distribution is sufficiently different from the expected distribution to conclude that the null hypothesis is incorrect and that there are true differences in the distribution of the populations (Crowl, 1993). Logic basic to the chi-square should inform the researcher that the larger the discrepancy between the actual number observed and the expected number in each category, the more likely the population values are not distributed proportionally. The larger the discrepancy, the larger the chi-square value will be, and the more likely it is that one will reject the hull hypothesis.

Analysis of a 1992 study by Sutarso shows that there was a statistically significant relationship between students' anxiety in learning statistics and the variables of student's achievement, statistical preknowledge, school, and current class

level. Sutarso detected some variables in relation to students' anxiety in learning statistics, but the results did not provide enough evidence to suggest that there was a relationship between student's anxiety in learning statistics and the other variables. Findings and conclusions such as this occur frequently in research (Thompson, 1988).

Further, consider a hypothetical study conducted to determine whether the proportion graduating from high school differs as a function of experimental condition, a null hypothesis was established:

$$H_o: \text{Graduate}_{Exp.} = \text{Graduate}_{Control}$$

The first step is to compute the expected frequency for each cell under the assumption that the null hypothesis is true as demonstrated in Table 3. The calculations for the χ^2 can then be computed yielding a $\chi^2 = 22.01$ has a probability value less than .0001. The results are found to be statistically significant, and the null hypothesis is rejected. However, to be useful to the researcher, a further analysis of the statistical breakdown must be accomplished to determine practical significance.

If the omnibus chi-square hypothesis is rejected, one would like to be able to find the contrast among the proportions that are significantly different from zero. Therefore chi-square analysis should not stop with the computation of an omnibus chi-

square statistic. Rather, additional post hoc tests are necessary.

Although the researcher may test any conceivable pair, only those related to the extreme values will actually result in a statistically significant difference. However, these values will not be present if the scores have been converted from continuous scores to nominal scores. Cox and Key (1993) suggested that these multiple pair-wise comparison tests enable the researcher to maintain the probability of experimentwise error at the prescribed value of alpha. Furthermore, these pair-wise comparisons also serve to identify possible causes for the rejection of the null hypotheses. When the overall analysis indicates that not all of the proportions are equal, the individual chi-square analysis may indicate differences that were statistically significant and were attributable to the rejection of the null hypothesis.

A major problem occurs in post-hoc tests, according to Thompson (1988), when multiple chi-square tests are performed in a single study and the test is applied to all possible pairs of variables. These tests can violate the validity of the chi-square tests since chi-square is based on the assumption that every observation is independent (Thompson, 1988). This type of research, in which the number of post-hoc tests escalates quickly,

$$\frac{\text{variables x (variables -1)}}{2},$$

has an impact on the increased probability of Type-I error. Although multivariate methods would be more appropriate under these conditions, the use of multiple chi-square tests is common, particularly in dissertation research (Stevens, 1996).

Beasley and Schumacher (1995) suggested that "it is possible to augment the omnibus and partitioned chi-square test by post hoc methods" (p. 89), arguing that a percentage of shared variance interpretation of chi-square results is needed. R^2 in multiple linear regression represents the proportion of variance shared between the dependent variable and the predicators. In the ANOVA this relation is explain by a categorical independent variable referred to as η^2 (eta-squared). An interpretation of shared variance is also possible in contingency chi-square tests.

Conclusion

Although technical advice was common in the first half of this century regarding the chi-square contingency table, commentaries offering direction for its use appear to be decreasing while the use of this form of statistical analysis continues to gain popularity.

In 1985, Rudolph, McDermontt, and Gold indicated that descriptive statistics, contingency tables, analysis of

variance, and t-tests were among the most commonly used statistical techniques. McCarney, in 1970, found that the predominant statistical techniques had changed over time in sociology from essentially descriptive statistics to more analytic methods such as correlation and chi-square. A 16 year extensive study (Emmons et al., 1990) reviewed articles from Sociology of Education, Journal of Education Psychology, and American Educational Research Journal. A surprise finding was that descriptive statistics, nonparametric techniques, and chi-square, usually associated with sociology, actually declined in use in Sociology of Education while increasing in use in the more quantitatively oriented American Educational Research Journal over the period of the study.

According to LaGaccia (1991,p. 153), "selection of inappropriate research methods can threaten the validity of conclusions made by researchers." To be so well known and so easily used, the chi-square contingency table statistic has been misused by educational researchers. While some researchers contend that the majority of variables analyzed by educational researchers are nominal or ordinal in nature, others suggest that the majority of variables are continuous in nature. It is incumbent upon researchers to know their data and to direct their research questions toward the most appropriate statistical procedures.

In completely analyzing the contingency table, the researcher has basically two options. One option is to calculate individual one-way chi-square values for each column of the table. The statistical significance can be reported separately for each level of the dependent variable. Another choice, Option II, involves a review of the contingency table and identifying differences in column percentages above a specified level. Option I is preferable when the statistically significant differences are detailed through specific explanations. Option II, which mainly involves a look and seek process, does not contain any test of statistical significance is of little value.

Wolfle (1980) offered that causal analysis with quantitative variable has become a useful means of understanding educational phenomena. Consequently, of the nonparametric tests of statistical significance, chi-square is the most frequently used by educational researchers in causal-comparative studies. With the advent of innovative statistical software programs such as SPSS and the ease of their use, the continued reliance on chi-square is a concern. More in-depth and detailed analyses of research data are available. Reluctance on the part of educators to incorporate technology into their methodology may attribute to the continued utilization of chi-square.

References

Beasley, T. M., & Schumacher, R. E. (1995). Multiple regression approach to analyzing contingency tables: Post hoc and planned comparison procedures. Journal of Experimental Education, 64, 79-93.

Best, J. W. (1981). Research in education (4th ed.). Englewood Cliffs: Prentice-Hall.

Burns, E.(1980). Applied research and statistics for teachers. Springfield, Illinois: Charles c. Thomas.

Chernoff, H., & Lehmann, E. L. (1954). The use of maximum likelihood in χ^2 tests for goodness of fit. Annals of Mathematical Statistics, 25, 579-586.

Coughlin, M. A., & Pagano, M. (1997). Case study application of statistics in institutional research. Resources in institutional research, number ten. Tallahassee, FL: Association for Institutional Research.

Cox, M. K., & Key. C. H. (1993). Post hoc pair-wise comparisons for the chi-square test of homogeneity of proportions. Educational and Psychological Measurement, 53, 951-62.

Crowl, T. K.(1996). Fundamentals of research (2nd ed.). Madison, WI: Brown & Benchmark.

Delucchi, K. (1981). The use and misuse of chi-square: Lewis and Burke revisited. Paper represented at the annual

meeting of the American Educational Research Association, Los Angeles, CA. (ERIC Document Reproduction Service No. ED 204 399)

Emmons, N. J., & others (1990). Statistical methods used in "American Educational Research Journal," "Journal of Educational Psychology" and "Sociology of Education" from 1972 through 1987. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA. (ERIC Document Reproduction Service No. ED 319 797)

Gall, M. D., Borg, W. R., & Gall, J. D. (1996). Educational research, an introduction (6th ed.). New York: Longman.

Goodwin, L. D., & Goodwin, W. L. (1985). Statistical techniques in AERJ articles, 1979-1983: The preparation of graduate students to read the educational research literature. Educational Researcher, 12, 5-11.

Kerlinger, F.N. (1986). Foundations of behavioral research (3rd ed.). New York: Holt, Rinehart and Winston.

LaGaccia, S. S. (1991). Methodology choices in a cohort of education dissertations. In B. Thompson (Ed.), Advances in educational research: Substantive findings, methodological developments (pp. 149-158). Greenwich, Conn.: JAI Press inc.

Lewis, D., & Burke, C. J. (1949). The use and misuse of the chi-square test. Psychological Bulletin, 46, 433-489.

Mouly, G.J. (1978). Educational research: The art and science of investigation. Boston: Allyn and Bacon.

Parshall, C. G., & Kromrey, J. D. (1996). Tests of independence in contingency tables with small samples: A comparison of statistical power. Educational and Psychological Measurement, 56, 26-44.

Popham, W. J., & Sirotnik, K. A. (1973). Educational statistics: Use and interpretation (2nd ed.). New York: Harper and Row.

Examining contingency table analysis results to detect assumption violations. Prophet StatGuide. Retrieved June 29, 1999 from the World Wide Web http://www.prophet.abtech.com/statguide

Stevens, J. (1996). Applied multivariate statistics for the social sciences (3rd ed.). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Sutarso, T. (1992). Some variables in relation to students' anxiety in learning statistics. Paper presented at the annual meting of the Mid-South Educational Research Association, Knoxville, TN. (ERIC Document Reproduction Service No. ED 353 334)

Thompson, B. (1988). Misuse of chi-square contingency-table test statistics. Educational and Psychological Research, 8, 39-49.

Thompson, B. (1991) (Ed.). Advances in educational research: Substantive findings, methodological developments. Greenwich, Connecticut: JAI Press inc.

Timm, N. H. (1971). Neyman's restricted chi-square tests. Paper presented at the annual meeting of the American Educational Research Association, New York, NY. (ERIC Document Reproduction Service No. ED 000 452)

Wolfle, L. M. (1980). Path analysis with categorical data: Applications to education. Paper presented at the annual meeting of the American Education Research Association, Boston, MA. (ERIC Document Reproduction Service No. ED 189 155)

Table I

Questions and Responses That Lead to a Chi-Square Analysis*

	Question	Response
a.	What are you testing	Differences
b.	What is the number and level measurement of the independent variable?	One independent variable Nominal level
c.	How many independent variables?	One independent variable
d.	How many levels or groups exist within the independent variable?	Two
e.	What is the type of independent variable?	Unmatched

*Note: Based in part on Coughlin and Pagano (1997).

Table 2

Calculation of 12 Participants' Throws

	Hit	Missed	Total
Preferred hand	3	9	12
Non-preferred hand	4	8	12
Total	7	17	24

Table 3

Is Graduating from High School a Function of Experimental Condition?

	Graduated	Failed to Graduate	Total
Exper.	73 (59.042)	12 (25.958)	85
Control	43 (56.958)	39 (25.042)	82
Total	116	51	167

$\chi^2 = 22.01$ $p < .0001$